2020 Census Redistricting Data (Public Law 94-171) Summary File
Development & Release Timeline

The 2020 Census Redistricting Data (Public Law 94-171) Summary File uses a powerful new disclosure avoidance system (2020 DAS) to protect against modern re-identification threats. It is the first 2020 Census data product to deploy the new system; the apportionment data, released April 26, 2020, remained unaltered by statistical noise.

We are developing customized applications of the 2020 DAS for the remaining 2020 Census data products. This timeline focuses on the key milestones in the development of the redistricting data product. Search "disclosure avoidance" at www.census.gov to learn more about our plans.

2019

Mar   (3/8) DSEP approves the proposed suite of 2020 Census data products.

(3/28) The Census Bureau releases the source code behind the prototype redistricting data based on 2018 Census Test data.

Apr   (4/15) USCB releases the prototype P.L. 94-171 (redistricting) file based on 2018 Census Test data using an early test version of the 2020 DAS.

Oct   (10/29) The Census Bureau releases first (baseline) 2010 Demonstration Data. This is the first of several releases that apply the latest iteration of the DAS to 2010 Census data for comparison purposes.

The DSEP-chosen privacy-loss budget (PLB) of 6.0 (4.0 for person records, 2.0 for housing) will be used on all development releases going forward to allow an "apples-to-apples" analysis of incremental DAS development progress. The DSEP will choose the final PLB after careful analysis, prior to final data processing.

Dec   (12/11-12) Data users provide invaluable feedback on the first demonstration data at a workshop hosted by the National Academies of Science's Committee on National Statistics (CNSTAT). Their feedback helps Census Bureau DAS developers identify additional use cases and informs ongoing development.

2020

Mar   (3/27) The Census Bureau releases initial quality metrics to allow easier analysis of ongoing DAS development progress for each demonstration data release. The Census Bureau also seeks input on additional measures to help data users assess each release and allow comparisons to published 2010 data.

May

(5/27) The Census Bureau releases second set of 2010 Demonstration Data and quality metrics. This produced a second round of detailed external user feedback including specific proposals for category binning and minimum-size geographic areas.

Aug

(8/26) Pandemic-triggered operational delays require the Census Bureau to shift the DAS development focus to attempt to meet the pre-pandemic data deadline. Focus shifts to redistricting data product-only until further notice. This limited the Census Bureau's ability to further develop the portions of the DAS that cover tables not included in the redistricting data product, including the ability to address many of the external suggestions from the first two demonstration data products.

Sept

(9/17) The Census Bureau releases third set of 2010 Demonstration Data and quality metrics, focused on redistricting data. This data product was based on the newly slimmed down code base for the redistricting data, and it contained several implementation errors that were discovered simultaneously by several external stakeholders and the internal development team.

Nov

(11/17) The Census Bureau releases fourth set of 2010 Demonstration Data and quality metrics, focused on redistricting data. This release corrected the implementation errors in the September demonstration data product and, once again, produced many analyses by external stakeholders that were used to refine the algorithms. In addition, this release was used to develop the final set of redistricting metrics, based on conversations with specialists in the redistricting community and at the Department of Justice.

(11/24) DSEP finalizes the list of "invariants" for the first set of 2020 Census data products. Invariants are statistics that are published without noise infusion.

Dec

The Census Bureau starts experiments to evaluate different settings of key system parameters for the redistricting data. Hundreds of full-scale experimental runs of the DAS will help the DSEP determine:
- The optimal set and processing order of queries against the confidential data.
- The share of PLB allocated to the different queries.
- The share of PLB allocated to tabulations at different geographic levels (e.g., county and tract).

2021

Mar

(3/25) DSEP approves the PLB for the fifth set of demonstration data. The PLB is set to ensure the accuracy of racial demographics for voting districts as small as 500 individuals. It is based on redistricting use cases informed by extensive

feedback from the redistricting community and the Civil Rights Division at the U.S. Department of Justice.

Apr (4/26) The Census Bureau releases 2020 Census apportionment data. Note that apportionment data are *not* affected by the DAS.

(4/28) The Census Bureau releases fifth set of 2010 Demonstration Data and quality metrics focused on redistricting data. These files used a higher PLB than previous releases that is more reflective of the anticipated final privacy/accuracy tradeoff for the 2020 Census data products. The released full privacy-loss accounting, based on the zero-Concentrated Differential Privacy (zCDP, discrete Gaussian mechanism), shows how the noise was injected. The optimized geographic spine used to improve accuracy for low-population minor civil divisions and places.

June (6/9) DSEP selects DAS parameters, including PLB, for redistricting data based on results of experimental runs and data user feedback.

Mid-June DAS development team implements parameters set by DSEP into algorithm for application to redistricting data product.

Late June Data production run and quality control analysis begins.

Aug (8/12) The Census Bureau provides a legacy format summary redistricting data file for each state.

The Census Bureau releases sixth and final set of 2010 Demonstration Data and quality metrics for the redistricting data. This "production-ready" set includes the final PLB and DAS parameters for the official 2020 Census redistricting data product.

Sept (9/16) The Census Bureau releases the redistricting data in user friendly formats such as data.census.gov.

(9/30) The Census Bureau releases the redistricting data source code—the algorithmic code and parameters that underpin the random assignment of noise infusion. The release enables experienced data scientists to analyze the TopDown Algorithm used to protect the redistricting data.

To learn more, search "Disclosure Avoidance" at www.census.gov.